

Bioinformatische Unterstützung der Auswahl von HIV-Therapien

Thomas Lengauer · André Altmann
Alexander Thielen

AIDS ist eine der wesentlichen Bedrohungen für die Menschheit unter den heutigen Infektionskrankheiten. Die Krankheit verursacht weltweit Millionen Todesfälle jährlich.

andere Bereiche der Welt, wie etwa das südliche Afrika, hoch mit AIDS durchseucht. Die Krankheit ist weltweit weiter auf dem Vormarsch.

Einführung

Der Erreger von AIDS, das menschliche Immunschwächevirus (Human Immunodeficiency Virus, HIV), ist vermutlich erst vor etwa hundert Jahren von Affen auf den Menschen übergesprungen. Der Erreger und sein neuer Wirt haben sich noch nicht co-evolutiv aneinander angepasst. Daraus resultiert die hohe Pathogenität des Virus beim Menschen – denn im Gegensatz zum Menschen werden Affen zwar von der ihnen entsprechenden Virusvariante (Simian Immunodeficiency Virus, SIV), mit der sie schon wesentlich länger leben, infiziert, entwickeln aber keine bedeutenden Krankheitssymptome.

AIDS ist aus verschiedenen Gründen besonders tückisch. Für die Bevölkerung ist besonders gefährlich, dass sich die Symptome der Krankheit nur über lange Zeiträume entwickeln, sodass ein Wirt lange Zeit infektiös sein kann, ohne dass er oder seine Kontaktpersonen es merken. Für den einzelnen Patienten ist besonders problematisch, dass das Virus sein Genom in das Genom der Wirtszelle einspleißt, sodass der Patient das Virus zeit seines Lebens nicht mehr los wird. Da, wie wir

Während in Europa aufgrund von Aufklärung und Vorbeugung relativ geringe Zahlen von Neuinfektionen zu verzeichnen sind (etwa 2700 im Jahr 2007 in Deutschland), sind

sehen werden, das Virus hoch variabel ist, ist es sehr schwierig, Impfstoffe dagegen zu entwickeln. Hier ist kein kurzfristiger Erfolg abzusehen. Da Heilung und Vorbeugung also stark behindert werden, zielen medikamentöse Therapien gegen AIDS darauf ab, die Virusreplikation im Patienten weitgehend zu unterdrücken und dadurch Symptome zu lindern und den Krankheitsverlauf hinauszuzögern.

Da sich das Virus im Patienten ständig verändert, haben die behandelnden Ärzte es mit einem bewegten Ziel zu tun. Als Antwort auf eine Wirkstofftherapie verändert sich das Virus – es wird resistent. Wirkstofftherapien müssen ständig auf das sich verändernde Virus im Patienten eingestellt werden. Die dabei wirkenden biologischen Zusammenhänge zwischen dem Virus-Genotyp und seiner Wirkstoffresistenz (dem Phänotyp) sind dabei hoch komplex und bis heute nicht grundlegend verstanden. Daher zieht man bei der Therapieauswahl im Wesentlichen Schlüsse aus in der klinischen Praxis gemachten Erfahrungen, die bis Anfang dieses Jahrhunderts ausschließlich manuell gesammelt und ausgewertet wurden. Die Komplexität der Zusammenhänge zwischen Genotyp und Phänotyp legt aber deren bioinformatische Analyse nahe. Wir haben zu diesem Zwecke den Server geno2pheno entwickelt, der solche Analysen frei auf dem Internet anbietet. Im Folgenden wird über die Ziele dieses Servers berichtet, die grundlegenden bioinformati-

DOI 10.1007/s00287-009-0352-7
© Springer-Verlag 2009

Thomas Lengauer · André Altmann · Alexander Thielen
Max-Planck-Institut für Informatik,
Campus E1 4,
66123 Saarbrücken, Deutschland
E-Mail: lengauer@mpi-inf.mpg.de

schen Methoden werden skizziert und der Einsatz des Servers wird beschrieben.

Vermehrungszyklus des Virus

Als Virus ist HIV keine autonom lebensfähige Einheit, sondern im Wesentlichen verpacktes Erbgut – knapp 10.000 genomische Buchstaben, die von einer Proteinhülle umgeben sind (Abb. 1). Dieser kurze genomische Text definiert einen der gefährlichsten bekannten biologischen Killer. Der Aufbau des HI-Virus ist heute sehr genau bekannt.

Wie alle Viren benutzt HIV die von ihm befallene Zelle – in der Regel eine Zelle des menschlichen Immunsystems –, um sich zu vermehren (Abb. 2). Die Kenntnis vom Vermehrungszyklus von HIV ist die Grundlage für alle Wirkstofftherapien gegen AIDS [11].

Das Genom von HIV besteht nicht aus dem allgemein bekannten Erbmolekül DNS, sondern aus der mit der DNS eng verwandten Ribonukleinsäure RNS, die beim Menschen als Überträgermolekül für Erbinformation verwendet wird. Das HI-Virus bindet mit seinem Hüllprotein (Abb. 1) zunächst an ein Oberflächenmolekül der zu infizierenden menschlichen Zelle und taucht dann in die Zelle ein. Wenn das Viruspartikel in die Zelle eingedrungen ist, wird seine Proteinhülle abgestreift. Dann übersetzt das Virus zunächst sein RNS-Genom in die von der menschlichen Zelle ablesbare DNS. Da die menschliche Zelle diese Aufgabe nicht übernehmen kann, hat das Virus ein spezifisch dafür entwickeltes Protein namens *Reverse Transkriptase* (RT, Rückübersetzer). Der Reversen Transkriptase fehlt ein Fehlerkorrekturmechanismus, sodass bei

praktisch jedem Kopiervorgang neue Varianten des HIV-Genoms entstehen. Das nach DNS übersetzte Virusgenom wird mithilfe eines weiteren vom Virus mitgebrachten Proteins, der *Integrase*, in das im Kern der befallenen Zelle vorhandene menschliche Genom eingebaut. Die normale Teilungsmaschinerie der menschlichen Wirtszelle ist damit gezwungen, auch das virale Genom zu vervielfältigen, und auch die Proteine des Virus werden von der Wirtszelle erzeugt. Die Komponenten des Virus sammeln sich in der Nähe der Zellperipherie. Neue Viruspartikel werden zusammengebaut und knospen von der Wirtszelle ab. Allerdings hängen mehrere virale Proteine in den unreifen Viruspartikeln noch zusammen. Sie müssen dadurch aktiviert werden, dass sie auseinander geschnitten werden. Das Auseinanderschneiden besorgt ein drittes Protein, das das Virus mitbringt, die sogenannte *HIV-Protease*. So kann jede befallene Immunzelle tausende Viruspartikel hervorbringen, bevor sie stirbt.

Entwicklung von Medikamenten gegen AIDS

Es gibt derzeit weit über 20 Medikamente gegen HIV auf dem Markt. Alle Medikamente sind kleine Moleküle und zielen darauf ab, einen Schritt im Lebenslauf von HIV zu blockieren, indem das Medikament an ein geeignetes funktionales Protein bindet. (Das durch das Medikament angegriffene Protein wird in diesem Zusammenhang *Zielprotein* genannt.) Es gibt mehrere Möglichkeiten, ein Protein zu blockieren. Eine beliebte Methode ist, den Wirkstoff so zu konzipieren, dass er sich an die

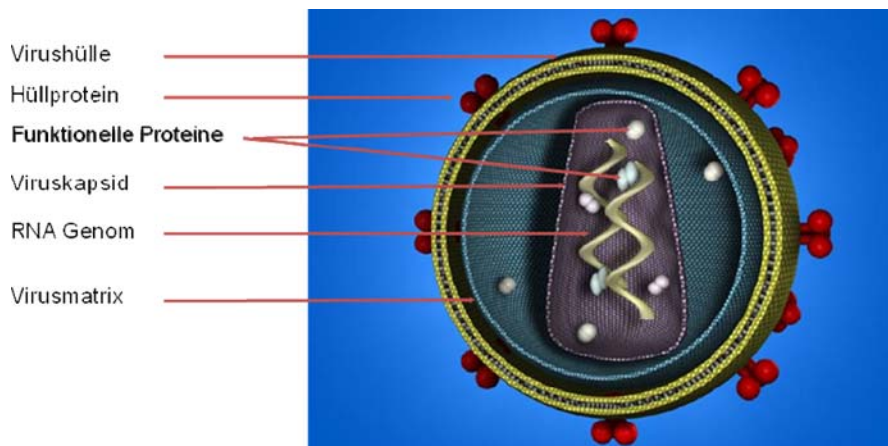


Abb. 1 Aufbau des Viruspartikels von HIV (Abbildung mit freundlicher Genehmigung von Michael Kurowski, Berlin)

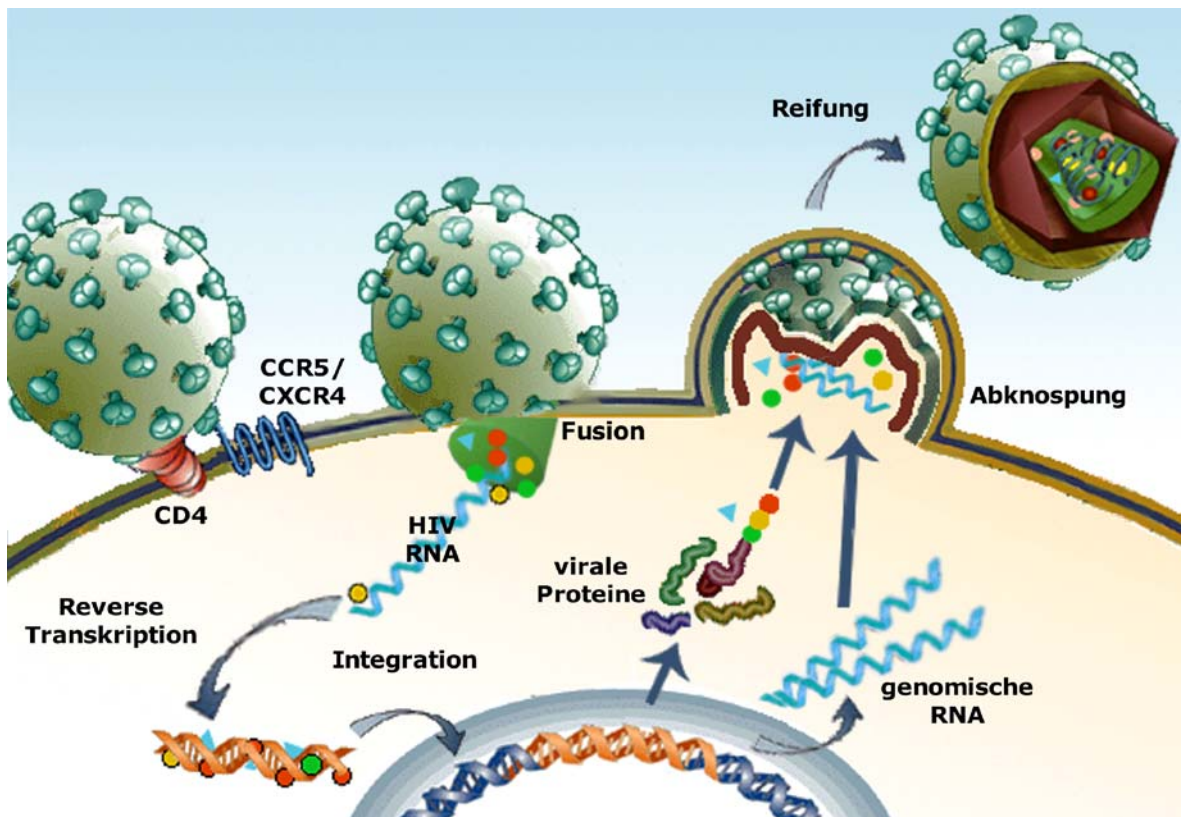


Abb. 2 Vermehrungszyklus von HIV, CD4 = zelluläres Membranprotein, an das das Virus beim Zelleintritt andockt, CCR5/CXCR4 = zelluläre Corezeptor-Proteine, die alternativ zusätzlich für den Zelleintritt benötigt werden

aktive Stelle am Protein heftet, also diejenige, an der das Protein an seinen molekularen Partner bindet. Zielproteine sind menschliche oder virale Hüllproteine zur Verhinderung des Zelleintritts des Virus sowie die drei Virusproteine Reverse Transkriptase, Integrase und Protease. Die weitaus größte Anzahl der gängigen AIDS-Medikamente greifen die virale Reverse Transkriptase und die Protease an.

Die Flucht des Virus in die Resistenz

Wäre das Virus nicht so variabel, würden ein oder zwei AIDS-Medikamente ausreichen. Das Virus verändert sich aber praktisch bei jedem Kopiervorgang. Die Baupläne der viralen Proteine sind im Virusgenom abgelegt und ändern sich, wenn das Genom mutiert. Die viralen Varianten, deren Proteine den Wirkstoff nicht oder weniger effektiv binden, haben dabei einen Selektionsvorteil und reichern sich im Patienten an: Das Virus wird resistent.

Je mehr Viren produziert werden, desto höher ist die Chance, dass resistente Formen entstehen und

sich anreichern. Gibt man parallel mehrere Wirkstoffe, die sich auch an verschiedene virale Proteine heften, so muss das Virus für eine erfolgreiche Flucht in die Resistenz eine größere Anzahl von Mutationen hervorbringen. Somit wird das Fortschreiten der Infektion verlangsamt.

Am Ende jedoch kann das Entstehen einer resistenten Virusvariante nicht verhindert werden. Dies passiert in der Regel im Verlauf von einigen Monaten bis zu ein oder zwei Jahren.

Herkömmliche Art der Therapieauswahl

Ist eine resistente Virusvariante entstanden, so steht der Arzt vor einer schwierigen Wahl, nämlich der einer geeigneten Therapie, die nun diese Variante unterdrückt. Gegen welche Wirkstoffe das Virus nun resistent ist, steht in seinem Genom geschrieben, jedoch in einer hoch verschlüsselten Form. Bis heute verstehen wir die molekularen Grundlagen der viralen Resistenz nicht. Daher haben sich zwei andere Zugänge zur Resistenzbestimmung herausgebildet.

Bei der sogenannten phänotypischen Resistenzanalyse [19] werden Viruspartikel in Zellkultur im Labor steigenden Konzentrationen einzelner Wirkstoffe ausgesetzt. Dabei wird beobachtet, wie sich die Vermehrungsrate des Virus reduziert. Aus dieser Beobachtung kann ein vergleichender Wert der Vermehrungsreduktion bei der viralen Variante aus dem Patienten und der Rate eines nicht resistenten Referenzvirus bestimmt werden, der sogenannte Resistenzfaktor. Ein Resistenzfaktor von 1 markiert die gleiche Resistenz des Virus vom Patienten wie des Referenzvirus. Hohe Resistenzfaktoren repräsentieren hoch resistente Viren. Ist der Resistenzfaktor z. B. 10, so braucht man die zehnfache Konzentration des Medikaments wie beim Referenzvirus, um eine Halbierung der Vermehrungsrate beim Virus des Patienten zu erreichen.

Phänotypische Daten haben mehrere Nachteile. Sie erfordern teure und langwierige Prozeduren, die nur in hoch spezialisierten Laboren durchgeführt werden können. Daher sind sie für die klinische Praxis relativ ungeeignet. Und sie bemessen die virale Resistenz in Zellkultur und nicht im Körper des Patienten. Sie liefern jedoch einen präzisen quantitativen Wert für das Ausmaß an viraler Resistenz.

In der klinischen Praxis geht man daher in der Regel anders vor. Man identifiziert die relevanten Teile des Genoms der im Patientenblut vorherrschenden viralen Variante, d. h., die Gene für die Zielproteine von AIDS-Medikamenten (genotypische Resistenzanalyse). Dies kann man schnell, kostengünstig und mit Routineprozeduren erledigen. Nun muss man jedoch die virale Resistenz (Phänotyp) aus dem Genotyp des Virus ablesen. Und gerade das ist sehr schwierig, da die Genotyp-Phänotyp-Beziehung hoch verschlüsselt und daher nicht einfach ablesbar ist. Man kann zwar klinische Parameter messen, die etwas über die Intensität der Infektion aussagen, sogenannte klinische Phänotypen. Dazu gehören die Viruslast, das heißt die Anzahl der Viruspartikel pro Milliliter Blut sowie die Anzahl der Immunzellen pro Mikroliter Blut ($CD4^+$ -Zellzahl). Je höher die Viruslast und je geringer die $CD4^+$ -Zellzahl, desto gravierender ist die Infektion. Diese Zahlen sind jedoch das Ergebnis der Reaktion des Virus auf mehrere Medikamente und auf das Immunsystem des Patienten. Aus ihnen ist also die Resistenz des Virus gegen einzelne Me-

dikamente nicht so klar abzulesen wie aus den im Labor gemessenen Resistenzfaktoren.

Bis vor kurzem hat man versucht, die Resistenz des Virus auf einzelne Medikamente manuell abzuleiten. Die klinischen Beobachtungen über Veränderungen im HIV-Genom und begleitende Resistenzen werden dabei in sogenannten Mutationstabellen gesammelt. Die am häufigsten verwendete derartige Tabelle wird von der International AIDS Society fortgeschrieben [9]. Die herkömmliche Art der Therapieauswahl ist nun wie folgt: Der Arzt sieht sich das aus dem Blut des Patienten isolierte Virusgenom an und sucht nach Wirkstoffen, gegen die keine der in der Mutationstabelle verzeichneten Resistenzmutationen zu finden sind. Eine Kombination solcher Wirkstoffe verabreicht er dann dem Patienten, wobei die genaue Auswahl vielfältige weitere Einflussfaktoren und Nebenbedingungen berücksichtigt.

Mutationstabellen haben einige schwerwiegende Nachteile. Zum einen geben sie vor, dass Mutationen im HIV-Genom unabhängig voneinander Resistenz verursachen. Dem ist aber nicht so. Es gibt Mutationen, die die Resistenz verstärken oder aufheben, die durch andere Mutationen verliehen wird [20]. Dieser expressive Nachteil von Mutationstabellen kann aufgehoben werden, wenn man computergestützte Expertensysteme benutzt. Diese können komplexere Regeln zur Resistenzbestimmung anwenden, die verschiedene Mutationen im HIV-Genom miteinander in Beziehung setzen. Solche Systeme sind heute auch vielfach in Gebrauch. Sie sind die moderne Version der Mutationstabellen [12, 16].

Es gibt jedoch noch einen zweiten Nachteil. Der manuelle Prozess zur Bestimmung von Resistenzmutationen (oder den entsprechenden Regeln) ist inhärent subjektiv und nicht reproduzierbar. Deswegen haben unsere virologischen Kooperationspartner um die Jahrtausendwende nach systematischeren Zugängen zur Resistenzanalyse gesucht. Die Bioinformatik bietet sich hier in natürlicher Weise als Methode an.

Bioinformatikmethoden zur Analyse von HIV-Resistenz

Die grundsätzliche Idee der bioinformatischen Resistenzanalyse ist, das Expertenwissen, das der manuellen Erstellung der Mutationstabellen oder der Regelsysteme zugrunde liegt, durch eine um-

fassende Datenbank über klinische Erfahrung mit AIDS-Therapien zu ersetzen. Zusammen mit unseren virologischen Partnern haben wir um die Jahrtausendwende mit dem Aufbau einer solchen Datenbank begonnen, die einen großen Anteil der in Deutschland verfügbaren Daten umfasst, der Arevir-Datenbank [13]. In den vergangenen drei Jahren haben wir die Datensammlung im Rahmen des EU-Projektes EuResist auf ein europäisches Niveau gehoben [14]. Die Arevir-Datenbank umfasst derzeit etwa 1000 Paare aus viralem Genotyp und Labor-Phänotyp. Die Euresist-Datenbank umfasst derzeit Daten zu knapp 100.000 Therapien von knapp 34.000 Patienten [21]. Dabei liegen für etwa 5400 Therapie-Episoden Genotypen und klinische Ergebnisse vor, d. h. der Zusammenhang zwischen viraler Sequenz und klinischem Erfolg ist explizit dokumentiert.

Auf der Basis solcher Daten lassen sich mit Methoden des maschinellen Lernens statistische Modelle für die Resistenzschätzung von neuen viralen Genotypen ableiten. Wir haben drei Arten von solchen Modellen erstellt.

Bestimmung von (neuen) Resistenzmutationen

Zum einen kann man mit statistischen Methoden den Informationsgehalt einer Mutation über die Resistenz des entsprechenden Virus bemessen. Dazu gibt es mehrere formale Möglichkeiten. Eine ist, zu beziffern, inwieweit sich resistente Viren in derjenigen Teilmenge der Datenbank anreichern, die nur aus Viren besteht, die die betreffende Mutation aufweisen. Ist dieser Anteil hoch, so ist auch der Informationsgehalt der Mutation bzgl. der Resistenz des Virus hoch, und sie kann als Resistenzmutation gelten. Auf diese Weise haben wir bis dato unbeachtete Resistenzmutationen gegen gängige AIDS-Medikamente gefunden, was von der Ärztesgemeinschaft mit großen Interesse aufgenommen wurde (Best Oral Presentation Award auf dem Europäischen HIV-Resistenzworkshop 2005) [18]. Diese univariate Analyse hat jedoch dieselben expressiven Nachteile wie die Mutationstabelle.

Resistenzschätzung auf der Basis des gesamten Virusgenoms

Die zweite Generation unserer Softwarewerkzeuge beinhaltet multivariate Analysen, die eine systematische Ableitung des in den manuell erstellten

Expertensystemen vorhandenen Wissens über die Genotyp-Phänotyp-Beziehung bei der viralen Resistenzanalyse ermöglichen. Wir haben eine ganze Reihe solcher Modelle erstellt. Dabei haben wir sowohl Verfahren zur Klassifikation des Virus (in die Kategorien resistent/nicht resistent) als auch zur regressiven Schätzung des Resistenzfaktors für ein Virus entwickelt. Für die Klassifikation setzen wir vor allem Entscheidungsbäume [2] und Random Forests ein [8], für die Regression Support-Vector-Maschinen [3]. Die statistischen Lernmethoden, die wir anwenden, sind state-of-the-art und werden an das jeweilige Problem angepasst. Im Zentrum unserer Arbeit steht die Modellierung, d. h. die Auswahl der Eingabe-Features und die geeignete Definition der Ausgabe (z. B. die Festsetzung der Definition der Kategorien resistent/nicht resistent bei der Klassifikation von Viren). Die Validierung der Methoden mit den daraus erwachsenden neuen medizinischen Einsichten über Resistenzentwicklung stellt einen weiteren wesentlichen Aspekt unserer Forschung dar.

Als Beispiel wird hier ein Entscheidungsbaum zur Klassifikation von HI-Viren bzgl. ihrer Resistenz gegen den Protease-Blocker Saquinavir erläutert (Abb. 3). Entscheidungsbäume sind hierarchische Strukturen, die die Entscheidungsfolgen repräsentieren, nach denen bestimmt wird, ob das Virus resistent gegen ein einzelnes Medikament ist. Die rechteckig eingekästelten Zahlen in Abb. 3 entsprechen Positionen in der Aminosäuresequenz des Zielproteins für das Medikament, in diesem Fall der Protease. Die durch einzelne Buchstaben angeordneten Aminosäurevarianten unter den eingekästelten Zahlen stellen mögliche Mutationen an der betreffenden Sequenzposition dar. Die Aminosäure des Referenzvirus ist mit einem grauen Buchstaben dargestellt. Graue Buchstaben repräsentieren also die Abwesenheit einer Mutation. Schwarze Buchstaben unter den eingekästelten Zahlen repräsentieren Abweichungen vom Referenzvirus, also Mutationen. An den unteren Enden der Entscheidungshierarchie befinden sich eingekreiste mit s (für suszeptibel = nicht resistent) markierte und mit r (für resistent) markierte Felder, die die endgültige Entscheidung über die Resistenz des Virus repräsentieren. (Die hier nicht näher erläuterten Zahlen unter diesen farbigen Feldern sagen etwas über die Genauigkeit der entsprechenden Vorhersage aus.) Der Baum in Abb. 3 liest sich wie folgt: Wenn an Sequenzposition

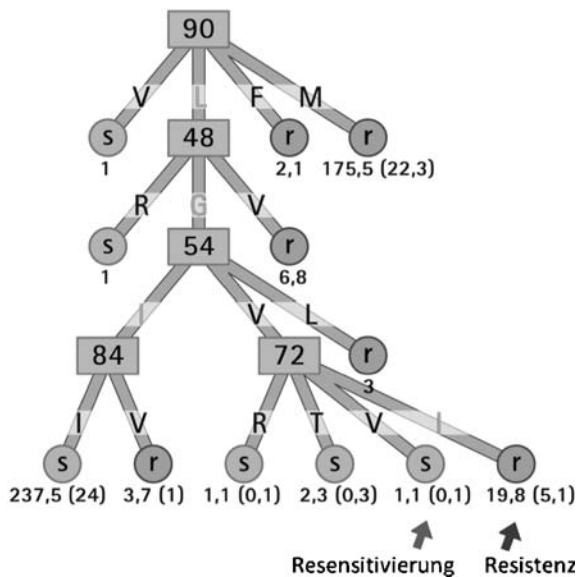


Abb. 3 Entscheidungsbaum für die Klassifikation eines Virus nach seiner Resistenz gegen das Protease-Medikament Saquinavir

90 der Protease eine Mutation von der (Aminosäure in der) Referenzvariante Leucin (L) nach Valin (V) zu finden ist, dann ist das Virus nicht resistent (gegen Saquinavir). Ist dort eine Mutation nach Phenylalanin (F) oder Methionin (M) zu finden, so ist das Virus resistent. Falls an dieser Position keine Mutation zu finden ist, müssen wir weiter an Position 48 schauen usw. Der Entscheidungsweg, der an der durch den Pfeil rechts unten markierten Stelle endet, zeigt, dass eine einzige Mutation an Position 54 von Isoleucin (I) nach Valin (V) zur Resistenz des Virus führt. Der Weg, der zur mit dem weiter links stehenden Pfeil markierten Stelle führt, zeigt jedoch, dass eine Mutation an Position 72 von Isoleucin (I) nach Valin (V) (oder auch Arginin (R) oder Threonin (T)) das Medikament wieder wirksam macht – sie hebt die durch die Mutation an Position 54 hervorgerufene Resistenz des Virus auf. Das ist ein Beispiel dafür, wie multivariate statistische Modelle die Abhängigkeiten zwischen Mutationen berücksichtigen.

In einem sogenannten Kreuzvalidierungstest stellte sich heraus, dass unsere Entscheidungsbäume in etwa 90% der Fälle korrekte Aussagen über die Resistenz eines Virus in der Zellkultur im Labor machen.

Unsere Resistenzmodelle für einzelne Wirkstoffe sind Basis des heute am meisten genutzten

Angebotes zur Resistenzanalyse auf dem geno2pheno-Server [23], den wir frei im Internet zur Verfügung stellen. Die behandelnden Ärzte oder Labors geben dort den Genotyp der im klinischen Kontext gefundenen viralen Variante ein und erhalten eine Übersicht wie in Abb. 4 gezeigt. Jede Zeile steht hier für ein Medikament. Die Medikamente im oberen Bereich der Abbildung blockieren die Reverse Transkriptase, die Medikamente im unteren Bereich blockieren die Protease. Die Spalte RF gibt die Schätzung des Resistenzfaktors (mittels Regression mit Support-Vector-Maschinen) an. Die Spalte z-score enthält Informationen über die Signifikanz des geschätzten Wertes für den Resistenzfaktor. Ist diese Zahl höher als 2, kann man von einer Resistenz des Virus gegen das Medikament ausgehen. Im rechten Teil werden Mutationen aufgelistet, und zwar in rot (hier hellgrau) solche, die Resistenz verstärken und in grün (hier dunkel) solche, die die Resistenz schwächen. Das geno2pheno Softwareangebot wird heute zur Therapie eines großen Teils (etwa zwei Drittel) der in Deutschland therapierten AIDS-Patienten genutzt [4, 10].

In unserer Forschung gehen wir jedoch weiter. Bei dem geno2pheno Angebot macht der Arzt nämlich immer noch sehr viel manuell. Zum einen muss er ausgehend vom geschätzten Resistenzfaktor, also einem Laborwert, auf die klinische Resistenz des Virus beim vorliegenden Patienten schließen. Zum anderen muss er verschiedene Wirkstoffe zusammenstellen, die auch miteinander interagieren können. Und schließlich ist nicht nur von Bedeutung, wie die gegenwärtige virale Variante auf die Therapie anspricht, sondern auch, wie lange ihr Fluchtweg in die Resistenz ist, wie lange die Therapie also wirksam bleibt. Diese Aspekte werden von der dritten Generation unserer Software zur Resistenzanalyse berücksichtigt.

Bioinformatikmethoden zur Auswahl von HIV-Kombinationstherapien

In Kombinationstherapien wirken Medikamente auf komplexe Weise zusammen. Daher kann man Resistenzen des Virus gegen einzelne gemeinsam verabreichte Medikamente nicht einfach maximieren oder addieren, um die Effektivität der Therapie zu schätzen. Um das Zusammenwirken von Medikamenten zu modellieren, benutzen wir deshalb wiederum statistische Lernverfahren, die subtile nichtlineare Abhängigkeiten zwischen den verschiede-

	Drug	RF	z-score	Scored Mutations
Medikamente gegen die Reverse Transkriptase	ZDV	87.9	7.9	215F 41L 70R 67N 49R
	ddC	2	3.8	184V 215F 74V 169D 219Q 70R 207E 135T
	ddI	3.8	5.7	74V 184V 215F 214F 218E
	d4T	1.8	2.8	67N 184V 215F 70R 169D 101N 41L 50V 207E 219Q
	3TC	99.7	16.8	184V 41L
	FTC	41.5	4.9	184V 202V 49R 215F 219Q
	ABC	9.6	13.6	184V 74V 41L 49R 215F 214F 181C 70R
	TDF	2.1	2.9	219Q 41L 70R 169D 184V 74V 60I 215F 67N 135T
	NVP	397.5	6.4	103N 181C 135T
	DLV	360	11.2	103N 181C 74V
	EFV	39.7	6.7	103N 181C 135T 214F
Medikamente gegen die Protease	SQV	1.4	0.9	63P 14R
	IDV	1.4	0.1	
	RTV	1.2	-0.1	63P 14R
	NFV	1.3	-0.1	63P 3I 12S
	APV	0.9	-0.6	63P
	LPV	1.1	0.2	63P
	TPV	1.2	-0.6	3I 63P 12S
	DRV	2.4	0.1	3I 37N 63P
	ATV	0.9	-1	14R 63P

Abb. 4 Ergebnis der Resistenzanalyse auf dem geno2pheno-Server

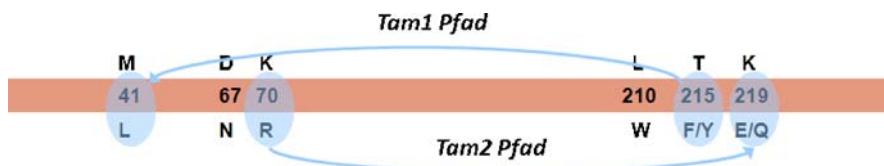


Abb. 5 Die ersten Mutationsschritte entlang zweier Fluchtpfade in die Resistenz gegen das AIDS-Medikament AZT

denen Wirkstoffen erkennen. Dabei betrachten wir nicht nur die Resistenz des gegenwärtig beobachteten Virusstammes, sondern wir extrapolieren auch in die Zukunft. Denn für den Erfolg einer Therapie ist nicht nur wichtig, wie resistent der gegenwärtige Virenstamm gegen die verabreichte Therapie ist, sondern auch, wie lange die Therapie gegen einen sich zunehmend verändernden Virenstamm effektiv bleibt.

Wie schon erwähnt, antwortet das Virus auf eine Wirkstofftherapie mit einer Flucht in die Resistenz – es setzen sich genomische Varianten durch, die gegenüber der Wirkstofftherapie Überlebensvorteile haben. Dabei ist zu beobachten, dass das Virus bestimmte Fluchtpfade in die Resistenz bevorzugt. Diese Pfade können heute noch nicht auf der Basis eines biologischen Verständnisses des Resistenzphänomens bestimmt werden. Man kann sie lediglich in der klinischen Praxis beobachten. So sind z. B. die in Abb. 5 dargestellten Fluchtpfade in die Resistenz gegen das Medikament AZT,

einen Reverse-Transkriptase-Blocker, beobachtet worden.

Im Prinzip sollten solche Fluchtpfade in einer Genotyp-Phänotyp-Datenbank zur HIV-Resistenz auffindbar sein. Dies ist dann besonders einfach, wenn sogenannte *longitudinale Daten* vorliegen. Dies sind Zeitreihen für einzelne Patienten, die im Verlaufe der Jahre therapeutisch begleitet werden, und bei denen jeweils über einen größeren Zeitraum Virustests vorgenommen wurden. Solche Zeitreihen stehen uns jedoch nicht in ausreichendem Maß zur Verfügung. Wir haben dagegen *cross-sektionale Daten* – also Daten von vielen Patienten, aber von jedem Patienten nur einen oder ganz wenige Datenpunkte. Aber auch in cross-sektionalen Daten sind Fluchtpfade auszumachen. Etwa gibt sich der TAM-1-Pfad dadurch preis, dass in unserer Datenbank eine ganze Reihe von Datenpunkten vorliegen, die Virenvarianten enthalten, die den TAM-1-Pfad mehr oder weniger weit durchschritten haben, also die Mutation an Position 215 (der Reversen

Transkriptase) oder aber die beiden Mutationen an den Positionen 215 und 41 aufweisen. Dagegen gibt es keine oder nur wenige Datenpunkte, die die Mutation an Position 41 haben, nicht aber die an Position 215. Wir haben eine Technik entworfen, mit der man eine bestimmte Art von statistischen (Bayes'schen) Modellen, sogenannte *Mutagenetische Bäume*, berechnen kann, die die Verteilung der Mutationsmuster in unserer Datenmenge möglichst genau wiedergeben [6]. Abbildung 6 zeigt das Ergebnis für die Fluchtpfade in die Resistenz gegen AZT. Der untere Baum bildet die beiden TAM-Pfade ab. Der obere Baum hat keine besondere Struktur und verkörpert das „Rauschen“ bei der Flucht in die Resistenz. Die Zahlen über den Bäumen besagen, dass 78% der Daten durch die beiden TAM-Pfade beschrieben werden, der Rest ist Rauschen.

Die Bäume werden vom Programm auch mit quantitativen Informationen versehen, die aussagen, in welchem Zeitraum die Resistenzmutationen zu erwarten sind. Mithilfe dieser Information kann man einen quantitativen Begriff für die sogenannte *genetische Barriere* ableiten [5]. Das heißt, man kann die Wahrscheinlichkeit beziffern, mit der das Virus als Antwort auf eine gegebene Therapie gegen einen gegebenen Wirkstoff in einem bestimmten Zeitraum ein bestimmtes Resistenzniveau erreicht. Die genetische Barriere ist die wesentliche Information, die es uns ermöglicht, die Effektivitätsdauer einer AIDS-Therapie abzuschätzen. Wir berechnen sie mithilfe der mutagenetischen Bäume und stellen sie einem umfassenden statistischen Modell zur Verfügung, das aus den Eingaben *vorherrschende virale Variante, frühere Therapien, geschätzte Resistenz der Variante gegen alle Wirkstoffe* und *genetische Barriere zur Flucht in die Resistenz gegen alle Wirkstoffe* den Erfolg einer gegebenen Kombinationstherapie von AIDS-Wirkstoffen abschätzt. Dies wird dann vergleichend für die mehreren hundert in der Praxis eingesetzten Kombinationstherapien vorgenommen. So werden die Therapien gemäß ihrer geschätzten Wirksamkeit gereiht. Solche Therapien, die die größten Chancen bieten, über einen gewissen Zeitraum wirksam zu sein, werden als erste vorgeschlagen. Der behandelnde Arzt sichtet die am besten bewerteten Therapien und berücksichtigt noch weitere Faktoren bei der Auswahl, die wir in unseren Modellen bisher nicht erfassen, wie etwa die Verträglich-

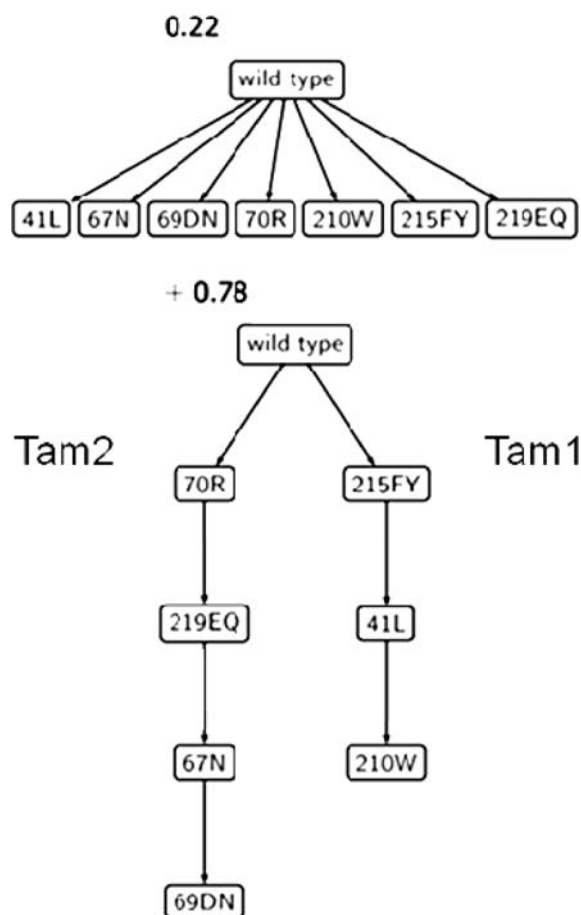


Abb. 6 Ein Modell mit zwei mutagenetischen Bäumen für die Flucht in die Resistenz gegen AZT. „Wild type“ bezeichnet hier den Ausgangspunkt für die Flucht in die Resistenz, nämlich das Referenzvirus

keit gewisser Medikamente durch den Patienten. Retrospektive Kreuzvalidierungsstudien haben gezeigt, dass unter Verwendung unserer Software die Fehlerrate bei der Auswahl von AIDS-Therapien von rund 1 von 4 (ohne Softwareunterstützung) auf 1 von 7 (mit unserer Software) reduziert werden kann [1].

Ein ermutigender Fall aus der klinischen Praxis

Für den Laien eindrucksvoller ist aber vielleicht folgende Geschichte aus der klinischen Praxis: Einer der mit uns zusammenarbeitenden Ärzte hat einen AIDS-Patienten in Behandlung, nennen wir ihn Georg. (Der wirkliche Name ist uns aus Datenschutzgründen korrekterweise nicht bekannt.) Georg hat sich im Jahre 1987 infiziert. Seitdem ist

er in Behandlung. Bisher hatte man aber keine Therapie gefunden, die ihm wirklich hilft. In den vergangenen 15 Jahren waren bei allen verabreichten Therapien wesentliche Zahlen von HI-Viren im Blut zu messen. Unabhängig davon durchlief Georg natürlich den Wechsel von Resistenzentwicklung und Therapiewechsel mehrmals. Im Oktober 2003 boten dann die Mutationstabellen keine Alternative mehr. Der im Blut von Georg isolierte Virenstamm hatte gegen alle Wirkstoffe Resistenzmutationen vorzuweisen. Nach der klassischen Vorgehensweise war Georg „austherapiert“. Der behandelnde Arzt zog zu diesem Zeitpunkt unseren Server zu Rate. Der Server machte einen Vorschlag, den der Arzt aufgrund seiner medizinischen Expertise noch etwas modifizierte. Seit diesem Therapiewechsel war Georg zum ersten Mal in seiner Patientenlaufbahn „virusfrei“, jedenfalls, was sein Blutserum betrifft. (Aus dem Körper kann man HIV, wie schon erwähnt, nicht ganz entfernen.) Dieser Zustand hält jetzt schon über fünf Jahre an.

Neue Medikamente

Auch noch so gute Software zur Therapieauswahl kann aber nicht verhindern, dass sich mit der Zeit im einzelnen Patienten, aber auch in der Bevölkerung allgemein, resistente Virenvarianten

anreichern. Daher ist es notwendig, immer neue AIDS-Medikamente zu entwickeln, die den Lebenslauf des Virus an immer anderen Stellen blockieren. Ein besonders attraktiver Punkt des Lebenszyklus ist in diesem Zusammenhang der Zelleintritt. Denn wenn man verhindert, dass das Virus in die Zelle eintritt, verhindert man auch seine Integration in das menschliche Genom und die komplette Elimination des Virus aus dem Körper des Patienten bliebe im Prinzip denkbar. Aus diesem Grunde hat sich die Forschung in den vergangenen Jahren intensiv auf die Entwicklung von Medikamenten zur Blockierung des Zelleintritts fokussiert, und erste solche Medikamente sind auch bereits auf dem Markt verfügbar und im klinischen Einsatz. Aber auch hier hat das Virus natürlich wieder evolutive Gegenmaßnahmen, um der Therapie zu entgehen.

Der virale Zelleintritt ist ein komplexer mehrstufiger Prozess, bei dem das Virus und die Wirtszelle wechselseitig miteinander wirken (Abb. 7).

Auf der Seite der Wirtszelle spielen dabei zwei Oberflächenproteine eine Rolle. Der zelluläre Rezeptor CD4 geht den ersten Kontakt mit dem Virus, genauer mit seinem Hüllprotein gp120, ein. In einem zweiten Schritt bindet ein weiteres zelluläres

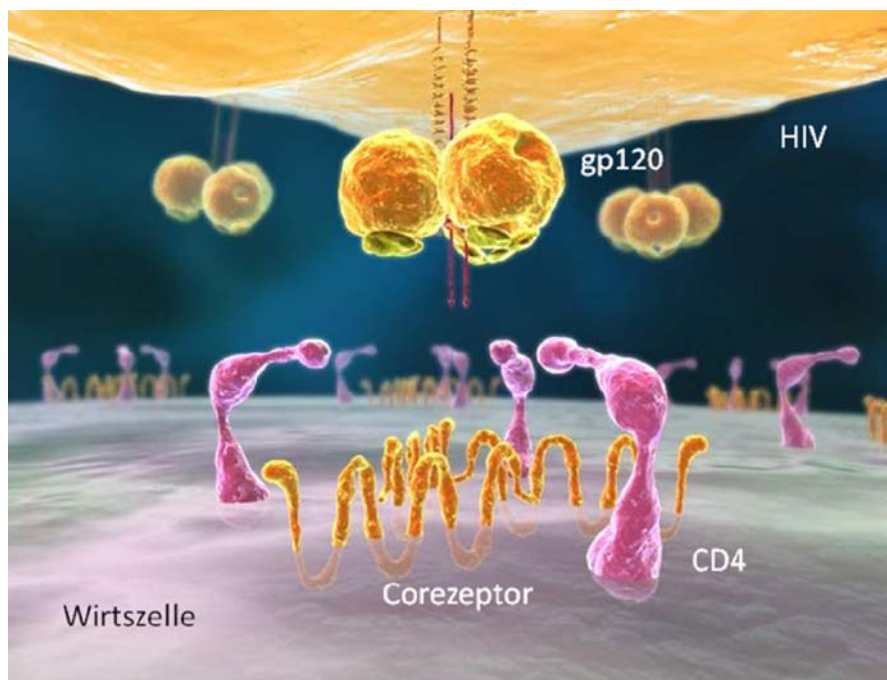


Abb. 7 Zelleintritt von HIV
(Mit freundlicher
Genehmigung von Pfizer
Pharma GmbH)

Oberflächenprotein, der sogenannte Korezeptor. Das Virus benutzt als Korezeptor eins von zwei verschiedenen zellulären Proteinen, die CCR5 und CXCR4 heißen. Eine Erstinfektion ist praktisch nur über den Korezeptor CCR5 möglich. Während der Infektion kann das Virus (durch Veränderung des Gens für sein Oberflächenprotein gp120) den Korezeptor wechseln und fortan weitere Zellen über den Korezeptor CXCR4 befallen. Ein solcher Korezeptorwechsel nach CXCR4 ist im Allgemeinen mit einem Fortschreiten der Krankheit assoziiert. Man versucht ihn daher zu vermeiden.

Neue Medikamente zum Blockieren des Zelleintritts greifen den Korezeptor CCR5 an. Dieses Protein ist ein attraktives Zielprotein für Medikamente. Man hat nämlich beobachtet, dass etwa 1% der Bevölkerung keine funktionelle Variante dieses Proteins haben. Solche Individuen können sich praktisch nicht mit HIV infizieren, haben aber sonst keine auffälligen Symptome. Der erste CCR5-Blocker kam im Oktober 2007 auf den Markt. Eine offensichtliche Antwort des Virus auf eine Therapie mit einem CCR5-Blocker ist der Wechsel des Korezeptors nach CXCR4. Daher ist bei einer Therapie mit CCR5-Blockern ein begleitender Test des Virus auf Korezeptorgebrauch zwingend vorgeschrieben. Für einen solchen gibt es – genau wie bei Resistenztests – eine kostspielige und aufwändige phänotypische Variante, bei der der Korezeptorgebrauch im Labor getestet wird. Wir bieten dafür auf dem geno2pheno-Server aber auch eine genotypische Variante an. Hierzu muss lediglich der relevante Abschnitt des viralen Hüllproteins gp120, der mit dem Korezeptor interagiert, in unseren geno2pheno-Server eingegeben werden. Unsere statistischen Modelle machen dann eine Schätzung für das Risiko, dass das Virus den Korezeptor CXCR4 benutzt. Dieses Serverangebot wird seit Herbst vergangenen Jahres besonders intensiv genutzt [15, 17].

Messung der viralen Quasispezies

Bei der Vorhersage des Korezeptorgebrauchs gibt es eine wesentliche Komplikation, die zwar im Prinzip auch bei der viralen Resistenz besteht, dort aber nicht dieselbe hohe klinische Bedeutung zu haben scheint. Diese besteht darin, dass im Körper des Patienten eine diverse Population verschiedener viraler Varianten vorherrscht, eine sogenannte Quasispezies. In unseren genotypischen Messungen erhalten wir in der Regel nur Aussagen über das Genom der

dominierenden viralen Variante. An Positionen im Genom, an denen zwei virale Varianten in großer Zahl (über 20% der Gesamtpopulation) vorkommen, erhalten wir einen sogenannten „Wobble“, d. h. eine zweideutige Information über den genomischen Text. Somit ist die gemessene virale Genomsequenz nur ein „Konsens“ der in der Population dominierenden Varianten. Genauere Informationen über den Aufbau der Quasispezies bleiben uns verborgen. Diese Tatsache beschränkt die Genauigkeit unserer Vorhersage zum Korezeptorgebrauch. So schaffen es unsere Modelle auf sogenannten *klonalen* Datensätzen, bei denen nur eine einzelne Virusvariante genotypisiert und phänotypisiert wurde, etwa 85% der CXCR4-gebrauchenden Viren zu erkennen, während bei klinischen Daten, in denen uns der Konsens über die dominierenden Varianten einer Quasispezies vorliegt, diese Genauigkeit auf etwa 67% abfällt. Dies resultiert zum einen aus der Konsensbildung beim Genotypisieren, vor allem jedoch dadurch, dass virale „Minoritäten“ – Varianten, die nur 10% oder weniger der Quasispezies ausmachen – überhaupt nicht aufgespürt werden.

In den vergangenen Jahren ist eine neue genomische Sequenzierungstechnologie entwickelt worden, das sogenannte *ultra-deep sequencing*. Damit ist es im Prinzip möglich, ein sehr genaues Abbild der gesamten viralen Quasispezies zu bekommen. Wir haben erste Erfahrungen mit dieser Technik gesammelt. Aus einer einzigen Probe Patientenblut haben wir dabei rund 125.000 Genabschnitte des viralen Hüllproteins gp120 identifiziert, die in über 1000 Varianten vorkamen. Wenn wir diese Varianten in unseren Vorhersageserver eingeben, erhalten wir ein Profil des Korezeptorgebrauchs der gesamten Quasispezies. Wir sind überzeugt, dass dieses Profil die notwendige Information ist, um unsere Vorhersagegenauigkeit bedeutend zu steigern. Validierungen dieser Technik stehen aber noch aus.

Ausblick

Trotz der Effektivität der entwickelten Methoden stehen wir im Hinblick auf eine adäquate Modellierung des Krankheitsgeschehens bei AIDS und der Wirkung von medikamentösen Therapien auf dieses Geschehen erst am Anfang. Eine wesentliche Auslassung besteht darin, dass das Genom des Patienten nicht in die Analyse einbezogen wird. Bei der Behandlung einer Infektionskrankheit gibt es

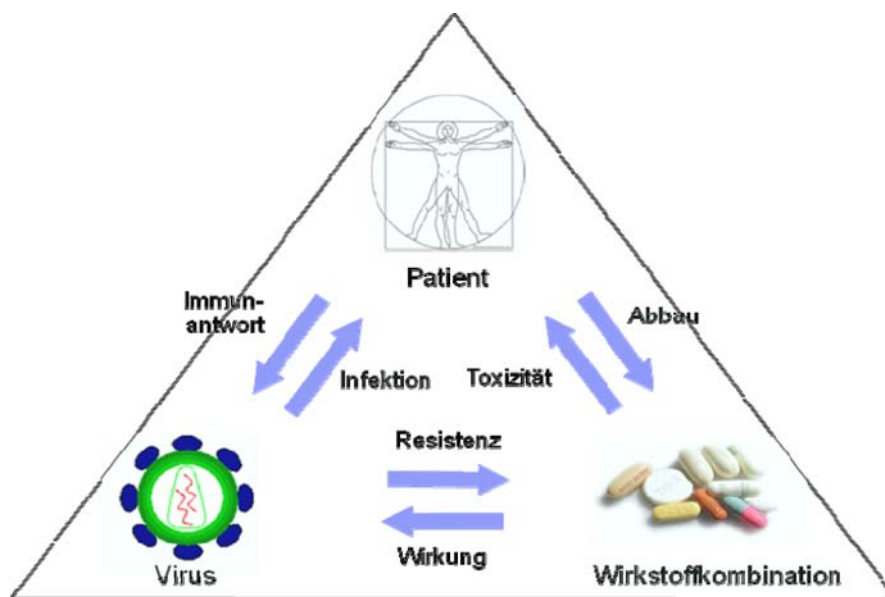


Abb. 8 Zusammenspiel der Akteure bei der HIV-Infektion

immer drei Akteure, den Patienten (Wirt), den Erreger (Gast) und den Wirkstoff (Abb. 8). Bisher haben wir nur die Beziehungen zwischen Erreger und Wirkstoff betrachtet. Es ist aber auch so, dass HIV auch Mutationen entwickelt, mit denen es versucht, dem Immunsystem des Wirtes zu entkommen.

Ein weiterer Ansatzpunkt für weiterführende Forschung ist, dass die der Resistenzanalyse unterliegenden klinischen Datenbanken trotz ihrer bereits beträchtlichen Größe für viele Medikamentenkombinationen nur ein geringes Datenaufkommen anbieten. Ferner sind die Datenbanken nicht repräsentativ für den heutigen globalen Wirtstoffeinsatz, sondern weisen regionale und zeitliche Verzerrungen auf. Es bedarf neuartiger statistischer Methoden, diesen Unzulänglichkeiten der Datenbasis entgegenzuwirken [7].

Darüber hinaus birgt das hier ausführlich besprochene therapeutische Ziel, die gegenwärtige Therapie möglichst lange effektiv zu halten, das Risiko, dass beim Versagen der Therapie – das unweigerlich früher oder später einsetzen wird – möglicherweise ein hoch resistenter Virenstamm vorliegt, der dann kaum weitere Therapieoptionen offen lässt. Daher sollte vermieden werden, durch die gegenwärtige Therapie die Anreicherung einer schwer therapierbaren viralen Variante zu provozieren. Diese Anforderung hat fundamentale Konsequenzen für den Therapieauswahlprozess.

Längerfristig ist es erstrebenswert, die bisher rein statistische Analyse klinischer Resistenzdaten

durch mechanistische Modelle der biochemischen Prozesse bei der Infektion von Zellen durch HIV anzureichern. Dabei findet der Infektionsprozess auf drei Ebenen statt: der Ebene der einzelnen befallenen Zelle, der Ebene des Organs im einzelnen Patienten sowie der Ebene der Patientenpopulation. Eine Multiskalensimulation der Krankheits- und Resistenzentwicklung auf allen drei Ebenen sowie ihrer Manipulation durch Wirtstofftherapien ist ein Forschungsziel, das noch in weiter Ferne steht.

Danksagung. Bioinformatische Resistenzanalyse ist ein interdisziplinäres Unterfangen. Die Urheber des Projektes waren vor neun Jahren Rolf Kaiser (jetzt Virologisches Institut der Universität zu Köln), Daniel Hoffmann (jetzt Universität Duisburg-Essen) und Joachim Selbig (jetzt Universität Potsdam). Besonders umfangreiche frühere Beiträge in unserem Labor kamen von Niko Beerenwinkel (jetzt ETH Zürich in Basel) und Tobias Sing (jetzt Novartis, Basel). Datensammlung in Deutschland sowie Validierung der Software und interdisziplinärer Dialog hinsichtlich neuer Fragestellungen werden von dem Verein genafor e. V. [22] getragen. Europaweite Datensammlung und Softwareentwicklung wird von dem Euresist Netzwerk [21] getragen. Wir danken besonders Martin Däumer, Richard Harrigan, Francesca Incardona, Michal Rosen-Zvi, Anders Sönnernborg, Hauke Walter und Maurizio Zazzi für die fruchtbare Zusammenarbeit. Das Projekt wurde

und wird von der DFG, dem BMBF und der EU finanziell unterstützt.

Literatur

1. Altmann A, Däumer M, Beerenwinkel N, Peres Y, Schülter E, Buch J, Rhee SY, Sönnernborg A, Fessel WJ, Shafer RW, Zazzi M, Kaiser R, Lengauer T (2009) Predicting the response to combination antiretroviral therapy: Retrospective validation of geno2pheno-THEO on a large clinical database. *J Infect Dis* 199(7):999–1006
2. Beerenwinkel N, Schmidt B, Walter H, Kaiser R, Lengauer T, Hoffmann D, Korn K, Selbig J (2002) Diversity and complexity of HIV-1 drug resistance: a bioinformatics approach to predicting phenotype from genotype. *Proc Natl Acad Sci USA* 99(12):8271–8276
3. Beerenwinkel N, Däumer M, Oette M, Korn K, Hoffmann D, Kaiser R, Lengauer T, Selbig J, Walter H (2003) Geno2pheno: estimating phenotypic drug resistance from HIV-1 genotypes. *Nucleic Acids Res* 31(13):3850–3855
4. Beerenwinkel N, Sing T, Lengauer T, Rahnenführer J, Roomp K, Savenkov I, Fischer R, Hoffmann D, Selbig J, Korn K, Walter H, Berg T, Braun P, Fätkenheuer G, Oette M, Rockstroh J, Kupfer B, Kaiser R, Däumer M (2005) Computational methods for the design of effective therapies against drug resistant HIV strains. *Bioinformatics* 21(21):3943–3950
5. Beerenwinkel N, Däumer M, Sing T, Rahnenführer J, Lengauer T, Selbig J, Hoffmann D, Kaiser R (2005) Estimating HIV evolutionary pathways and the genetic barrier to drug resistance. *J Infect Dis* 191(11):1953–1960
6. Beerenwinkel N, Rahnenführer J, Däumer M, Hoffmann D, Kaiser R, Selbig J, Lengauer T (2005) Learning multiple evolutionary pathways from cross-sectional data. *J Comput Biol* 12(6):584–598
7. Bickel S, Bogojeska J, Lengauer T, Scheffer T (2008) Multi-task learning for HIV therapy screening. In: McCallum A, Rowe S (eds) 25th International Conference on Machine Learning (ICML 2008). Omnipress, Helsinki, pp 56–63
8. Breiman L (2001) Random forests. *Mach Learn* 45(1):5–32
9. Johnson VA, Brun-Vezinet F, Clotet B, Gunthard HF, Kuritzkes DR, Pillay D, Schapiro JM, Richman DD (2008) Update of the drug resistance mutations in HIV-1: Spring 2008. *Top HIV Med* 16(1):62–68
10. Lengauer T, Sing T (2006) Bioinformatics-assisted anti-HIV therapy. *Nat Rev Microbiol* 4(10):790–797
11. Modrow S, Falke D, Truyen U (2003) *Molekulare Virologie*, 2. Aufl. Spektrum Akademischer Verlag, Heidelberg Berlin
12. Rhee SY, Gonzales MJ, Kantor R, Betts BJ, Ravela J, Shafer RW (2003) Human immunodeficiency virus reverse transcriptase and protease sequence database. *Nucleic Acids Res* 31(1):298–303
13. Roomp K, Beerenwinkel N, Sing T, Schülter E, Büch J, Sierra-Aragon S, Däumer M, Hoffmann D, Kaiser R, Lengauer T, Selbig J (2006) Arevir: A secure platform for designing personalized antiretroviral therapies against HIV. In: Third International Workshop on Data Integration in the Life Sciences (DILS 2006). Springer, Hinxton
14. Rosen-Zvi M, Altmann A, Prosperi M, Aharoni E, Neuvirth H, Sönnernborg A, Schülter E, Struck D, Peres Y, Incardona F, Kaiser R, Zazzi M, Lengauer T (2008) Selecting anti-HIV therapies based on a variety of genomic and clinical factors. *Bioinformatics* 24(13):399–406
15. Sander O, Sing T, Sommer I, Low AJ, Cheung PK, Harrigan PR, Lengauer T, Domingues FS (2007) Structural descriptors of gp120 V3 loop for the prediction of HIV-1 coreceptor usage. *PLoS Comput Biol* 3(3):e58
16. Sing T, Däumer M (2006) Interpretation algorithms. In: Geretti M (ed) *Antiretroviral resistance in clinical practice*. Medscape, London, pp 43–56
17. Sing T, Low AJ, Beerenwinkel N, Sander O, Cheung PK, Domingues FS, Büch J, Däumer M, Kaiser R, Lengauer T, Harrigan PR (2007) Predicting HIV coreceptor usage on the basis of genetic and clinical covariates. *Antiviral Therapy* 12(7):1097–1106
18. Svicher V, Sing T, Santoro MM, Forbici F, Rodriguez-Barrios F, Bertoli A, Beerenwinkel N, Bellocchi MC, Gago F, d'Arminio Monforte A, Antinori A, Lengauer T, Ceccherini-Silberstein F, Perno CF (2006) Involvement of novel human immunodeficiency virus type 1 reverse transcriptase mutations in the regulation of resistance to nucleoside inhibitors. *J Virol* 80(14):7186–7198
19. Walter H, Schmidt B, Korn K, Vandamme AM, Harrer T, Überla K (1999) Rapid, phenotypic HIV-1 drug sensitivity assay for protease and reverse transcriptase inhibitors. *J Clin Virol* 13(1–2):71–80
20. Wolf K, Walter H, Beerenwinkel N, Keulen W, Kaiser R, Hoffmann D, Lengauer T, Selbig J, Vandamme AM, Korn K, Schmidt B (2003) Tenofovir resistance and re-sensitization. *Antimicrob Agents Chemother* 47(11):3478–3484
21. <http://www.euresist.org>, letzter Aufruf: 18.5.2009
22. <http://www.genafor.de>, letzter Aufruf: 18.5.2009
23. <http://www.geno2pheno.org>, letzter Aufruf: 18.5.2009